

Iterative Development of XR Learning Designs for Music Education Through Design-Based Research: Evidence from Teacher Professional Development

Philipp Ahner

University of Music Trossingen
Trossingen, Germany
p.ahner@doz.hfm-trossingen.de

Tobias Rotsch

University of Music Trossingen
Trossingen, Germany
t.rotsch@doz.hfm-trossingen.de

Benedict Saurbier

Osnabrück University
Osnabrück, Germany
benedict.saurbier@uni-osnabrueck.de

Leonard Bruns

Osnabrück University
Osnabrück, Germany
lebruns@uni-osnabrueck.de

Christian Geiger

University of Applied Sciences Düsseldorf
Düsseldorf, Germany
geiger@hs-duesseldorf.de

Julia Lückhof

University of Applied Sciences Düsseldorf
Düsseldorf, Germany
julia.lueckhof@hs-duesseldorf.de

Sandro Dalfovo

University of Music Trossingen
Trossingen, Germany
s.dalfovo@doz.hfm-trossingen.de

Tray Minh Voong

Osnabrück University
Osnabrück, Germany
tray.minh.voong@uni-osnabrueck.de

Michael Oehler

Osnabrück University
Osnabrück, Germany
michael.oehler@uos.de

Abstract—Extended Reality (XR) offers new possibilities for music education, but empirically evaluated XR learning designs and corresponding teacher professional development formats remain scarce. This study examines the iterative development of three XR learning designs within the LEVIKO-XR project using a Design-Based Research approach. The designs were implemented and revised across six teacher professional development workshops in Germany with 64 participants. Questionnaire data were analyzed using cycle-specific descriptive statistics, linear trend models, and Spearman rank correlations. Teachers' evaluations of the XR learning designs generally became more positive across successive DBR cycles. The strongest and most consistent development was observed for PatchWorld — a collaborative environment in which learners invent and build music together in a shared virtual space — particularly regarding intention to use the design, connection to music teaching, low-threshold access, and student and teacher motivation. Professional development quality also developed positively, with the largest increase in practice relevance and high later-cycle ratings for structure and exchange. These findings are consistent with increasingly positive evaluations across iterative implementation, evaluation, and revision. Because different participant groups attended the workshops and no controlled longitudinal design was used, the trends are interpreted as indicators of iterative development rather than as causal effects.

Index Terms—Virtual Reality, Mixed Reality, Design-Based Research, LEVIKO-XR

I. INTRODUCTION

Extended Reality (XR) technologies are increasingly discussed as educational media, but their specific potential for

music education remains underexplored. Existing research has addressed music-related XR and virtual environments from different perspectives [1], including neuroscientific approaches to music experience [2], coordination and motor learning in musicians [3], and rhythm- or game-based learning contexts [4]. At the same time, first music education concepts have begun to investigate how XR technologies can be used in music classrooms and music-related learning environments [5]–[8]. However, empirically evaluated XR learning designs for music education, particularly in the context of teacher professional development, remain scarce.

The implementation of digital music technologies presents a particular educational challenge because devices, software applications, user interfaces, and regulatory conditions change continuously. Teaching and learning scenarios therefore cannot simply be transferred unchanged from established musical practices to new technological environments [9], [10]. Technological literacy and technology acceptance should instead be understood as dynamic prerequisites that require continuous development [11]. For music teachers, this creates a need for professional development formats that provide opportunities for experimentation, practical musical experience, and reflection on the relationship between musical content, pedagogical objectives, and technological affordances.

Professional development for digital music education should therefore enable teachers to experience technologies from a learner's perspective and to reflect on the interaction of content, pedagogy, and technology [12]. At the same time, music education should avoid technologically deterministic perspectives. The educational value of XR does not arise from

This work was supported by the European Union (EU) and the Federal Ministry for Education, Family Affairs, Senior Citizens, Women and Youth (BMBFSFJ).

the technology alone, but from the ways in which teachers and learners appropriate it for musical and pedagogical purposes [13]. The central question is thus not whether XR is inherently beneficial, but how pedagogically meaningful learning designs can be developed, implemented, evaluated, and revised in authentic educational contexts.

Design-Based Research (DBR) offers a suitable framework for addressing this challenge because it connects the development of educational interventions with their repeated implementation and evaluation in practice [14]. Rather than treating an initial design as a finished product, DBR uses successive cycles to refine educational materials, instructional structures, and implementation conditions. This appears particularly relevant for XR-based music education, where technological change, usability, didactic accessibility, and institutional conditions are closely intertwined.

The LEVIKO-XR project addresses this research gap by developing, testing, and evaluating adaptable XR learning designs for music education. The collaborative project combines perspectives from music education, music psychology, musicology, and XR development. Previous project publications have addressed individual learning designs, XR-based musical practices, social presence, musical interaction, and implementation in music education [15]–[19]. The present study complements these contributions by focusing specifically on the quantitative development of teachers’ evaluations across successive DBR cycles.

The study examines whether iterative revisions of the XR learning designs and the accompanying professional development format were associated with changes in teachers’ evaluations. Particular attention is paid to the perceived didactic potential of the learning designs, their relevance for music teaching, and the quality of the professional development activities. The following research questions guide the analysis:

- 1) How did teachers’ evaluations of the XR learning designs develop across successive DBR cycles?
- 2) Which XR learning designs and item-level dimensions showed the strongest positive trends across cycles?
- 3) How did teachers’ evaluations of professional development quality develop across the DBR cycles?

II. METHODS

A. Project Context and Design-Based Research Approach

LEVIKO-XR was conducted as a collaborative research and development project at the intersection of music education, music psychology, musicology, and XR development. The project followed a Design-Based Research approach in which educational designs were developed, implemented in authentic professional development settings, evaluated, and revised across successive cycles [14]. Each cycle combined the implementation of the XR learning designs with formative reflection on technical, didactic, and organizational aspects.

B. Didactic Design and XR Learning Designs

Rather than developing bespoke software, the project deliberately grounded its learning designs in applications that were

already available, including commercial products. Accordingly, the object of development and iterative revision was not the underlying XR software itself, but the pedagogical learning design around these applications, including the selection and configuration of content where applicable, task sequences, guidance and scaffolding, social arrangements, and their integration into the professional development format. This ensured that teachers and learners engaged with mature, robust tools and could relate the experience to everyday musical practice. The approach widened the range of accessible content but also imposed limits: in some areas—among them music theory and music history—no available application met the quality criteria the project had established. To span as much of this content space as possible, three learning designs with distinct thematic emphases were created, and these form the core of the professional development modules. Throughout this paper they are denoted by the applications underpinning them: *Rhythm Games* (LEVIKO1), *Soapbox* (LEVIKO2, virtual music performances in a mixed-reality classroom), and *PatchWorld* (LEVIKO3, collaborative music invention).

Table I, adapted from [20], gives, for each design, two illustrative learning activities that condense the underlying subject-didactic content into selective, highly compressed examples, together with the knowledge domains each design addresses in relation to the SEPACK.digital competence model [21]. At the level of the professional development itself, a key intended outcome is that teachers become able to adapt these designs to the conditions of their own practice.

C. Teacher Professional Development Cycles

The learning designs were implemented in six teacher professional development workshops in Germany with a total of 64 participants. One workshop was exploratory and not part of the systematically evaluated series. The exploratory workshop served to pilot the workshop structure and technical setup and informed revisions implemented before the first systematically evaluated cycle. The five subsequent workshops were evaluated with the questionnaire and constitute the five evaluated DBR cycles (C1–C5) analyzed in this study. Following each workshop, questionnaire data, internal reflection documents, observations, and group discussions were reviewed by the project team and informed revisions of both the learning designs and the professional development format before the subsequent cycle.

The evaluated modules lasted approximately 1.5 hours and followed a comparable core structure: a brief introduction and onboarding phase, three application-specific modules, and a concluding reflection. Participants rotated through all modules, with the module order varying across rotation groups, and worked predominantly in pairs or small groups. A total of 14 XR headsets were available, resulting in approximately 30–45 minutes of direct XR exposure per participant. The modules were facilitated by project researchers, and questionnaire data were collected immediately after each module and at the end of the workshop.

TABLE I
SUBJECT-DIDACTIC TEACHING–LEARNING-DESIGNS IN LEVIKO-XR.

Teaching–learning-design		LEVIKO1: VR Rhythm Games — experiencing, understanding, and reflecting on rhythms and metres in virtual environments [22]	LEVIKO2: Virtual music performances in a mixed-reality classroom	LEVIKO3: Collaborative music invention with mood clips in the XR application PatchWorld [15]
Example activity 1	learning	Comparative reflection on one’s own movements across different levels of various VR rhythm games	Discussing aesthetic forms of musical expression (performances) using one’s own examples and different mixed-reality performances	Negotiating and modulating aesthetic decisions (sound parameters) to design sounds in the virtual “sound cinema” world
Example activity 2	learning	Playing different levels in the VR rhythm game <i>Moon Rider</i> with conscious perception of music, rhythm, metre, and holistic body movement	Social exchange about one’s own experiences and preferences regarding the performances addressed, and about the emotional effects of music (emotion-wheel diagram)	Collaboratively producing a short original piece of music and an accompanying video
Example knowledge domains (SEPACK.digital)		Subject-matter knowledge (popular-music rhythm); pedagogical content knowledge (on the sustainable acquisition of rhythmic competencies)	Subject-matter knowledge (aesthetic forms of expression in pop music); knowledge of the subject dimension of disciplinary education in the digital transformation (learners’ everyday worlds)	Knowledge of the object dimension of disciplinary education in the digital transformation (virtual objects, ways of making and producing music)

Note. The table lists two representative examples of central learning activities per design; they sketch the underlying subject-didactic content in a selective and strongly condensed form. The bottom row indicates examples of the connection to the SEPACK.digital competence model.

The nature and extent of the revisions differed across the three learning designs. Tasks and interaction in VR Rhythm Games (LEVIKO1) were only marginally refined, with the primary focus on clarifying how the design could be implemented didactically in music lessons. The design was further developed by strengthening its connection to music-didactic knowledge and competencies related to rhythm learning and by enhancing social interaction through tandem work, with one participant using VR and the other following the streamed view on a tablet. Key concrete changes included modifications to the social formats used when inventing movements (individual work within the whole group in C1; pair work in C2; and pair and small-group work followed by whole-group interaction in C3 and C4), the addition of didactic input in the form of teacher-led instruction, which was progressively developed from C1 to C5, changes to the amount of playing time and the use of individual levels depending on the available time (longer in C3 and C4 than in C1 and C2, but shorter in C5 for reasons of time management), and an increased emphasis on social communication within the pairs in relation to embodiment from C2 onward.

In LEVIKO2, the content was revised more substantially, likewise with the aim of strengthening participating teachers’ music-didactic understanding of the learning content, namely musical performances in Soapbox, and of increasing participant activation. This included, for example, more specific analytical tasks related to the performances (implemented and refined from C3 onward), the establishment of aesthetic comparisons with participants’ familiar or favourite artists, and expanded communication within the practical phases through the use of an online board (in C4, skipped in C5 for reasons of time management). Further changes comprised an expansion of the MR exploration space to outdoor areas (in C3 and C4), the use of multiple media formats (YouTube videos and the 2D version of Soapbox on smartphones from C3 onward), and

the introduction of paper-based worksheet tasks in C3. These tasks were complemented by a paired exchange activity and a brief whole-group discussion at the beginning of the module, focusing on the elements to be explored. The focus and scope of these tasks were subsequently refined from C3 to C5.

In PatchWorld (LEVIKO3), the thematic focus of the VR workshop sessions was continuously developed over the course of the design process. For instance, the proportion of time devoted to exploratory phases in the created virtual sound-cinema environment was reduced (through a range of measures, such as the introduction of a one-page learning sheet in C3, the addition of an onboarding presentation as a teacher-led instructional unit from C3 onward, and an increased focus on this aspect by the facilitating instructors from C4 onward), while more concrete work-related and music-making tasks were introduced in order to guide participants toward the intended learning activities. Moreover, in C4 and C5, priority was placed on product orientation, namely the completion and presentation of a joint composition as a screen recording, which was associated with a further (and observed) increase in participants’ motivation. In addition, new music-pedagogical tasks aimed at fostering musical interaction were formulated and progressively refined from C2 to C5 (in C4 and C5 partly reduced).

The underlying XR applications themselves remained unchanged.

D. Evaluation Instruments and Data Corpus

The broader project evaluation combined formative and summative elements. In line with the DBR approach [14], internal evaluation documents, semi-structured observation protocols, and group discussion protocols were used to reflect on the individual workshop modules and to inform subsequent design decisions. These qualitative materials were part of the formative DBR process. However, they are not systematically

analyzed in the present paper. Instead, the present contribution focuses on questionnaire data in order to examine quantitative evaluation trends across DBR cycles.

The quantitative data comprise questionnaire-based ratings of the perceived educational potential of XR technologies, the didactic potential of the individual XR learning designs, and the quality of the teacher professional development workshops. The learning-design questionnaire addressed three XR-based units: Soapbox, PatchWorld, and Rhythm Games. Items covered perceived student motivation, teacher motivation, connection to music teaching, low-threshold access, and intention to use the learning design in music lessons. Four learning-design items were available for all three XR learning designs: perceived student motivation, teacher motivation, connection to music teaching, and low-threshold access. An additional item on intention to use the learning design in music lessons could only be analyzed for PatchWorld. Although analogous intention-to-use items were included for the other learning designs, they could not be evaluated because the corresponding datasets were incomplete. Accordingly, this item is reported only for PatchWorld and is not included as a common comparison dimension across all three learning designs. Responses were given on a seven-point Likert-type scale, with higher values indicating more positive evaluations.

The professional development quality items were based on the Fortbildungsmonitor by Richter and Richter [23]. They covered four dimensions of process quality: structure, practice relevance, cognitive activation, and exchange. These items were rated on a four-point Likert-type scale, with higher values indicating more positive evaluations. Because the exploratory workshop was not part of the evaluated series, the professional development quality scale was analyzed for the five evaluated workshops (DBR cycles 1–5).

E. Statistical Analysis

The analysis was designed to examine whether the iterative DBR cycles were accompanied by systematic changes in the evaluation of the XR learning designs and the professional development format. First, item responses were cleaned and recoded so that higher values consistently represented more positive evaluations. For each participant, construct scores were calculated as arithmetic means of the corresponding items. Construct-level scores were computed for the overall evaluation of XR learning designs, for the individual XR applications, and for the professional development quality dimensions.

For each construct, descriptive statistics were calculated separately for each DBR cycle, including means, standard deviations, valid sample sizes, and 95% confidence intervals. These descriptive results were used to visualize the development of the learning-design evaluations and professional development ratings across cycles.

To assess trends across DBR cycles, the workshop cycles were treated as an ordered variable. For each construct and each individual item, a simple linear trend model was estimated with the evaluation score as dependent variable and

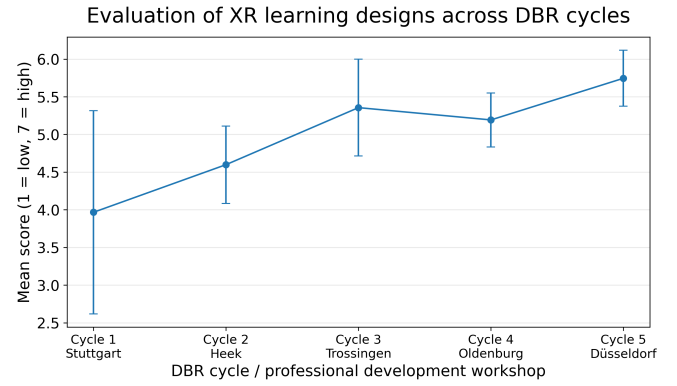


Fig. 1. Overall evaluation of XR learning designs across the DBR cycles. Points represent mean scores across evaluated XR learning design items; error bars indicate 95% confidence intervals. Higher values indicate more positive ratings on the 1–7 response scale.

DBR cycle as predictor. The unstandardized slope coefficient b indicates the average change in scale points per DBR cycle. In addition, Spearman’s rank correlation ρ was calculated as a non-parametric measure of the monotonic association between DBR cycle and item or construct score. This was included because the data were ordinal, the sample sizes per cycle were small, and the study was conducted in a naturalistic professional development context rather than as a controlled experiment.

Item-level trend analyses were used exploratively to identify which specific evaluation items showed the strongest positive development across cycles. The strongest positive item-level trends are reported in Table III and visualized in Fig. 4. Given the exploratory DBR context, the statistical analyses are interpreted as indicators of iterative development rather than as causal evidence. No correction for multiple testing was applied to the item-level analyses, as these were used for descriptive and exploratory purposes.

III. RESULTS

The evaluation data indicate a generally positive development of the XR learning designs across the DBR cycles. As shown in Fig. 1, the overall rating of the XR learning designs increased from the first to the fifth cycle. The mean score rose from approximately 4.0 in Cycle 1 to approximately 5.8 in Cycle 5 on the 1–7 scale. Although the confidence intervals indicate variation between cycles, the overall pattern suggests that the iterative DBR process was accompanied by increasingly positive evaluations of the learning designs.

A more differentiated view by application is provided in Fig. 2. The most pronounced increase was observed for PatchWorld, whose mean ratings rose from approximately 3.2 in Cycle 1 to above 5.8 in Cycle 5. Rhythm Games received comparatively high ratings already in the earlier cycles and remained at a high level across the later cycles. Soapbox showed positive evaluations in the cycles in which it was assessed, although its trajectory was less continuous because no Cycle 5 value is available due to incomplete questionnaire

TABLE II
ITEM-LEVEL DEVELOPMENT OF XR LEARNING DESIGN EVALUATIONS ACROSS DBR CYCLES.

Design	Item	Short item wording	C1	C2	C3	C4	C5	<i>b</i>	<i>p</i>	ρ
Soapbox	DI04_01	Student motivation	4.60	4.73	5.78	5.00	–	0.19	.372	0.06
Soapbox	DI04_02	Teacher motivation	4.40	4.36	5.56	4.36	–	0.07	.750	0.02
Soapbox	DI04_03	Connection to music teaching	4.20	4.55	5.44	4.82	–	0.23	.298	0.12
Soapbox	DI04_04	Low-threshold access	5.20	4.73	5.78	5.09	–	0.09	.691	0.07
PatchWorld	DI02_01	Student motivation	3.33	3.91	5.33	4.91	5.20	0.49	.011	0.37
PatchWorld	DI02_02	Teacher motivation	3.00	3.73	4.44	4.91	5.80	0.66	< .001	0.48
PatchWorld	DI02_03	Connection to music teaching	3.33	4.00	5.25	5.09	6.20	0.66	< .001	0.53
PatchWorld	DI02_04	Low-threshold access	3.67	4.00	5.56	4.82	6.00	0.52	.005	0.41
PatchWorld	DI02_05	Intention to use in music lessons	2.67	3.45	5.56	5.55	6.00	0.91	< .001	0.55
Rhythm Games	DI03_01	Student motivation	4.40	5.55	5.44	5.82	5.20	0.18	.275	0.11
Rhythm Games	DI03_02	Teacher motivation	4.40	4.82	5.11	5.64	5.40	0.31	.101	0.18
Rhythm Games	DI03_03	Connection to music teaching	5.20	5.45	4.78	5.45	5.60	0.06	.680	-0.02
Rhythm Games	DI03_04	Low-threshold access	5.20	5.73	5.25	5.91	6.40	0.23	.220	0.09

Note. C1–C5 denote the five DBR cycles. Cycle columns report mean ratings on a 1–7 scale, with higher values indicating more positive evaluations. Missing values indicate that the respective learning design was not evaluated in that cycle. *b* denotes the unstandardized linear slope per DBR cycle. ρ denotes Spearman's rank correlation between DBR cycle and item score. The *p* values refer to the linear trend tests. Short item labels are used for readability; the full translated item wordings are provided in Table IV. The intention-to-use item was analyzable only for PatchWorld because the corresponding data for Soapbox and Rhythm Games were incomplete.

TABLE III
STRONGEST POSITIVE ITEM-LEVEL TRENDS ACROSS DBR CYCLES.

Item	Area	Short description	<i>b</i>	<i>p</i>	ρ	<i>n</i>
DI02_05	PatchWorld	Intention to use in music lessons	0.91	< .001	0.55	42
DI02_02	PatchWorld	Teacher motivation	0.66	< .001	0.48	42
DI02_03	PatchWorld	Connection to music teaching	0.66	< .001	0.53	41
DI02_04	PatchWorld	Low-threshold access	0.52	.005	0.41	42
DI02_01	PatchWorld	Student motivation	0.49	.011	0.37	42
UX01_01	PatchWorld UX	Possibilities of PatchWorld meet requirements	0.46	.044	0.35	36
RI02_03	Training	Addressing problems from everyday professional life	0.41	< .001	0.55	40
RI03_01	Training	Incorporation of prior knowledge	0.35	< .001	0.51	41
RI01_01	Training	Clearly communicated objectives	0.35	< .001	0.51	40
DI03_02	Rhythm Games	Teacher motivation	0.31	.101	0.18	41
RI02_02	Training	Relation to current professional practice	0.29	.003	0.41	41
RI02_04	Training	Applicability to professional practice	0.29	.005	0.38	41

Note. *b* denotes the unstandardized linear slope per DBR cycle. ρ denotes Spearman's rank correlation between DBR cycle and item score. Items are sorted by descending positive slope. The *p* values refer to the linear trend tests.

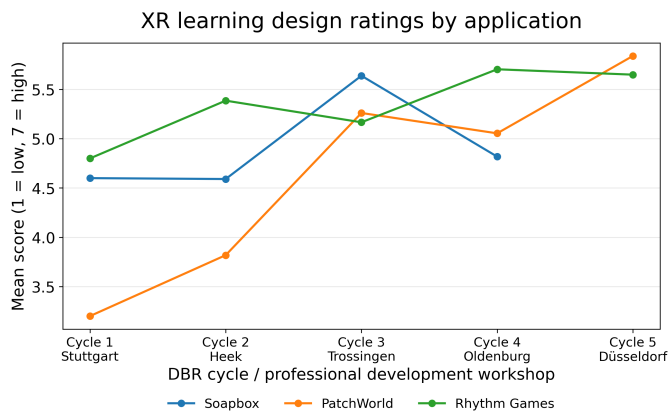


Fig. 2. Evaluation of XR learning designs by application across DBR cycles. The figure shows mean ratings for Soapbox, PatchWorld, and Rhythm Games on the 1–7 response scale. Missing lines indicate cycles in which the respective application was not evaluated.

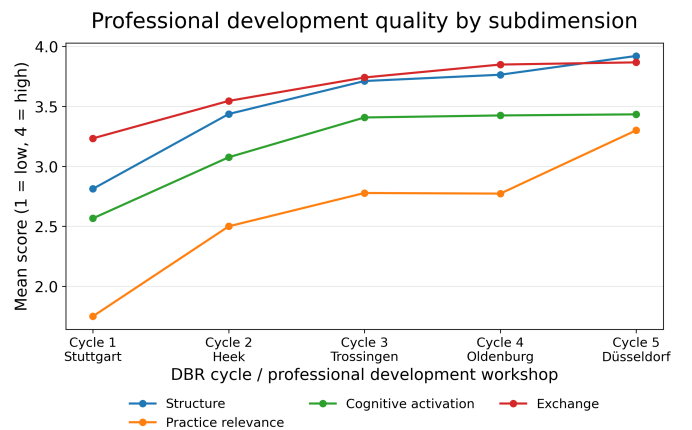


Fig. 3. Professional development quality by subdimension across DBR cycles. Mean scores are shown for structure, practice relevance, cognitive activation, and exchange. Higher values indicate more positive ratings on the 1–4 response scale.

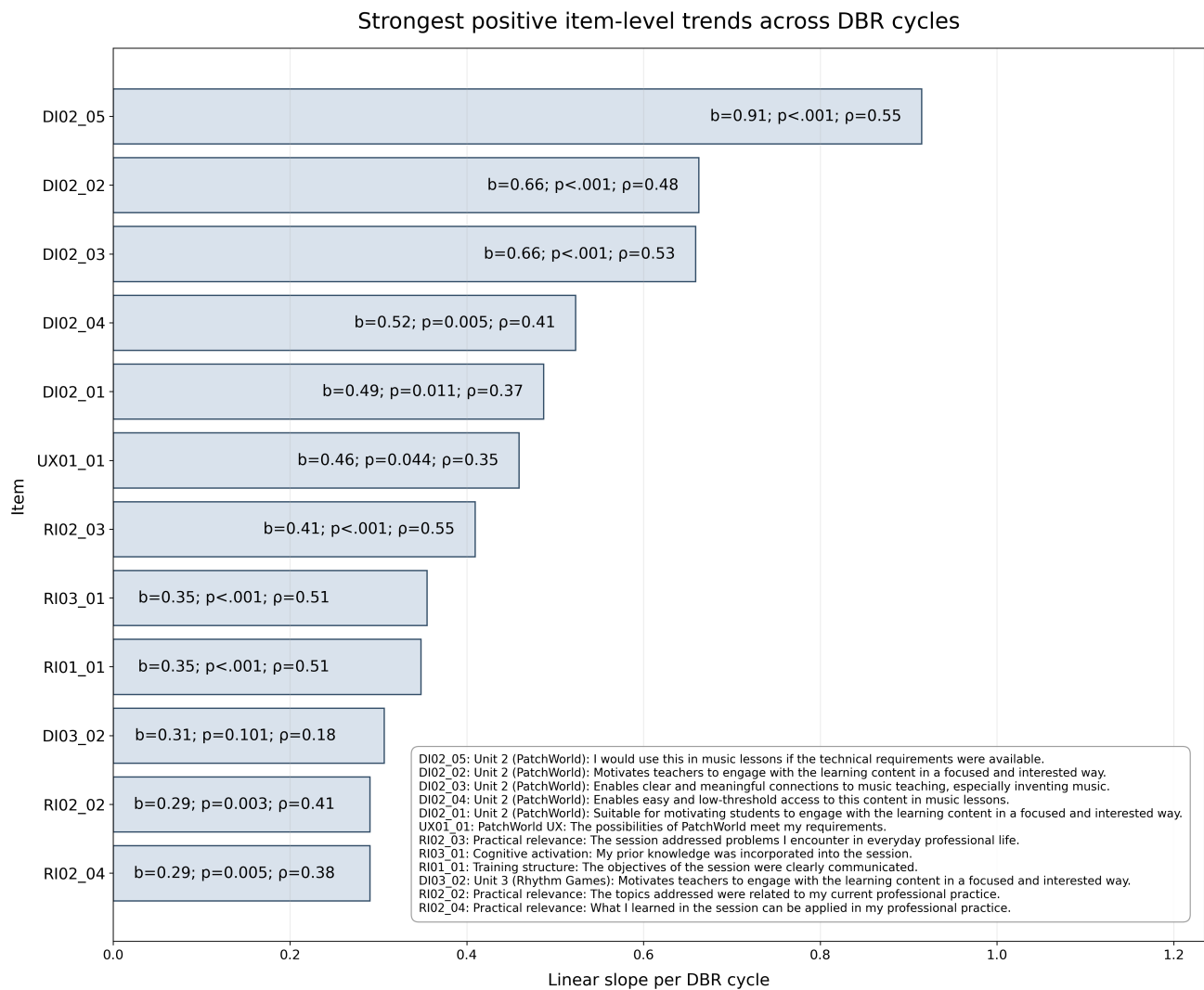


Fig. 4. Strongest positive item-level trends across DBR cycles. Bars show the unstandardized linear slope per DBR cycle. In-bar annotations report the slope coefficient b , the corresponding trend-test p value, and Spearman's rank correlation ρ between DBR cycle and item score. The text box resolves item codes into English item descriptions.

data. Overall, the application-specific results suggest that the iterative refinement was particularly visible for PatchWorld, while other applications showed more stable or less complete trajectories.

Table II provides an item-level view of the three XR learning designs and helps explain the aggregated application-level trajectories shown in Fig. 2. Short item labels are used for readability; the full translated item wordings are provided in Table IV. The clearest and most consistent development was observed for PatchWorld, where all five items increased from the early to the later DBR cycles. In particular, the intention to use PatchWorld in music lessons showed the strongest positive trend ($b = 0.91$, $p < .001$, $\rho = .55$). Soapbox and Rhythm Games started from comparatively higher or more stable ratings, but showed weaker item-level trends. This suggests that the aggregated increase for PatchWorld was not driven by a single item, but reflected improvements across

several didactically relevant dimensions.

The professional development evaluation also showed positive developments across cycles. Fig. 3 displays the development of four quality dimensions: structure, practice relevance, cognitive activation, and exchange. All four dimensions increased from the first to the later cycles. The largest improvement was observed for practice relevance, which started at the lowest level in Cycle 1 and increased substantially by Cycle 5. Structure and exchange reached high ratings in the later cycles, approaching the upper end of the 1–4 response scale. Cognitive activation also increased and stabilized at a positive level from Cycle 3 onward. These findings suggest that not only the XR learning designs, but also the professional development format itself was refined over the course of the project.

To identify which specific items contributed most strongly to these developments, item-level trend analyses were conducted. Fig. 4 shows the strongest positive item-level slopes across

DBR cycles, and Table III provides the corresponding statistics. The strongest trends were concentrated in the PatchWorld-related design items. The largest increase was found for teachers' intention to use PatchWorld in music lessons, with an unstandardized slope of $b = 0.91$ per DBR cycle, $p < .001$, and a moderate positive rank correlation of $\rho = .55$. Other PatchWorld items also showed substantial positive trends, including teacher motivation ($b = 0.66$, $p < .001$, $\rho = .48$), perceived connections to music teaching ($b = 0.66$, $p < .001$, $\rho = .53$), low-threshold access ($b = 0.52$, $p = .005$, $\rho = .41$), and student motivation ($b = 0.49$, $p = .011$, $\rho = .37$).

In addition to the application-specific findings, several professional development items showed positive item-level trends. These included the extent to which the training addressed problems from everyday professional practice ($b = 0.41$, $p < .001$, $\rho = .55$), the incorporation of participants' prior knowledge ($b = 0.35$, $p < .001$, $\rho = .51$), and clearly communicated objectives ($b = 0.35$, $p < .001$, $\rho = .51$). These results indicate that positive developments across the iterative DBR process were observed both for the XR learning designs and for the quality of the accompanying professional development activities.

IV. DISCUSSION

The findings suggest that the iterative DBR process was accompanied by increasingly positive evaluations of both the XR learning designs and the professional development format. The most pronounced development was observed for PatchWorld, whereas Soapbox and Rhythm Games showed comparatively stable trajectories. These differences are discussed below in relation to iterative design refinement, accessibility, practice

relevance, and the conditions of teacher professional development.

Regarding the first research question, the overall ratings indicate that the XR learning designs became increasingly convincing to participating teachers across the successive cycles. The development was not entirely linear, as illustrated by the slight decrease in the overall score in Cycle 4, but the general trajectory was positive. This pattern is consistent with the DBR principle of repeatedly implementing, evaluating, and revising educational interventions in authentic settings [14]. Rather than treating the first implementation as a finished product, the workshops functioned as opportunities to identify technical, organizational, and (subject) didactic limitations and to adapt subsequent versions.

With regard to the second research question, PatchWorld showed the strongest development, whereas Soapbox and Rhythm Games started from comparatively higher ratings and exhibited weaker trends. Part of this difference may reflect PatchWorld's lower initial ratings, which left greater room for upward movement; regression to the mean and chance differences between participant cohorts therefore cannot be excluded as alternative explanations for the steeper observed slope. At the same time, the extent of refinement differed between the designs: Rhythm Games were only marginally revised, whereas PatchWorld underwent more continuous changes toward concrete and product-oriented music-making tasks, while the Soapbox trajectory lacks a final-cycle evaluation. Because DBR cycle, workshop location, participant cohort, and project stage were confounded, the stronger PatchWorld trend cannot be attributed uniquely to the revisions themselves. The findings therefore do not establish PatchWorld as inherently superior to the other applications, but indicate that its development was

TABLE IV
FULL ITEM WORDING FOR THE XR LEARNING DESIGN EVALUATION ITEMS.

Design	Item	Full item wording
Soapbox	DI04_01	This learning design is suitable for motivating students to engage with the learning content in a focused and interested way.
Soapbox	DI04_02	This learning design motivates teachers to engage with the learning content in a focused and interested way.
Soapbox	DI04_03	This learning design made possible connections to music teaching clearly visible.
Soapbox	DI04_04	This learning design enables easy and low-threshold access to the respective content in music lessons.
PatchWorld	DI02_01	This learning design is suitable for motivating students to engage with the learning content in a focused and interested way.
PatchWorld	DI02_02	This learning design motivates teachers to engage with the learning content in a focused and interested way.
PatchWorld	DI02_03	This learning design made possible connections to music teaching, especially inventing music, clearly visible.
PatchWorld	DI02_04	This learning design enables easy and low-threshold access to the respective content in music lessons.
PatchWorld	DI02_05	I would use this learning design in music lessons if the technical requirements were available.
Rhythm Games	DI03_01	This learning design is suitable for motivating students to engage with the learning content in a focused and interested way.
Rhythm Games	DI03_02	This learning design motivates teachers to engage with the learning content in a focused and interested way.
Rhythm Games	DI03_03	This learning design made possible connections to music teaching, especially rhythm and movement, clearly visible.
Rhythm Games	DI03_04	This learning design enables easy and low-threshold access to the respective content in music lessons.

Note. The item wordings are English translations of the German questionnaire items. All items were rated on a 1–7 scale, with higher values indicating more positive evaluations.

particularly visible across the observed cycles.

The strongest PatchWorld trend concerned teachers' intention to use the learning design if the necessary technical conditions were available. This suggests that acceptance was not limited to general interest in XR but extended to a prospective willingness to integrate the design into teaching practice. At the same time, the conditional wording of the item points to the continuing importance of infrastructure, hardware availability, institutional support, and reliable technical conditions. Positive evaluations of a learning design therefore do not automatically imply that it can be implemented sustainably in everyday school practice.

The findings also support a technology-specific but human-centered understanding of digital music education. New devices cannot simply replace established musical tools while instructional methods remain unchanged [10]. At the same time, educational change should not be attributed to technology itself, but to the ways in which teachers and learners combine technological affordances with musical goals and pedagogical practices [13]. In the present project, the learning designs were revised not only with regard to software operation, but also in relation to learner activity, musical objectives, classroom accessibility, and the role of the teacher. DBR appears particularly suitable for this context because it allows educational designs to remain responsive to changing software, interfaces, and institutional conditions without making technological novelty the primary educational objective [9]. This emphasis on the pedagogical appropriation of XR technologies also connects the present findings to broader work on music in extended realities and emerging musical-metaverse practices [1], [8].

Regarding the third research question, the evaluation of professional development quality also developed positively. The strongest increase was observed for practice relevance, while structure and exchange reached values close to the upper end of the scale in the later cycles. This is important because teachers' acceptance of a learning design depends not only on the perceived quality of the application, but also on whether the professional development format enables them to relate the experience to their own teaching contexts. Professional development for digital music education should therefore provide spaces for practical experimentation and professional exchange while encouraging reflection on the relationship between musical content, pedagogy, and technology [12].

The experiential structure of the workshops may have contributed to these evaluations. Teachers did not merely receive information about the applications, but participated in practical modules similar to those intended for students. Such an approach can support reflection on both learner experiences and pedagogical decisions. It may also help participants deal with uncertainty associated with unfamiliar technologies. In rapidly changing digital environments, technical difficulties cannot always be eliminated. Professional development may therefore benefit from framing glitches and errors as opportunities for collaborative problem solving rather than solely as failures [12]. However, the present questionnaire data did not

directly measure changes in teachers' technical self-efficacy or tolerance of errors, and this interpretation should therefore be examined in future research.

The learning designs consisted of comparatively focused musical units. Such short and accessible formats can support initial engagement with unfamiliar technologies, but they should not be mistaken for complete models of long-term musical development. Music education discourse has raised concerns that isolated short-term activities may conflict with broader and more holistic understandings of musicianship [24]. A continuing challenge is therefore to connect motivating introductory experiences with sustained musical learning processes [25]. The positive evaluations in this study indicate that the learning designs can provide productive points of entry, but further research is required to examine how they can be embedded in longer curricular sequences.

The observed developments primarily concern the micro level of specific workshops and teaching-learning situations. Whether they translate into curricular change at the meso level or institutional change at the macro level cannot be determined from the present data [26], [27]. Sustainable implementation will likely require not only well-designed learning activities, but also access to equipment, technical support, curricular integration, organizational commitment, and continued opportunities for professional exchange.

Several limitations must be considered. First, the workshops involved different participant groups. The observed trends therefore do not represent longitudinal changes within the same individuals. DBR cycle, workshop location, participant composition, and project stage were confounded, making causal attribution impossible. Second, the sample sizes were comparatively small and varied between items and cycles. Third, some applications and professional development dimensions were not evaluated in all workshops, limiting comparability. Fourth, the analyses relied on self-reported evaluations rather than observed classroom implementation, student learning outcomes, or sustained technology use.

In addition, the item-level analyses were exploratory and involved multiple statistical tests without correction for multiplicity. Individual p values should therefore not be interpreted as confirmatory evidence. The convergence between descriptive trajectories, linear slopes, and rank correlations is more informative than any isolated significance test. Finally, although qualitative documents, observations, and group discussions informed the formative DBR process, they were not systematically analyzed in the present paper. Integrating these materials with the quantitative findings could provide a more detailed account of which revisions were made between cycles and why particular changes were effective.

Future research should follow teachers and students longitudinally, examine classroom transfer after professional development, and investigate whether positive evaluations lead to sustained use. Controlled or matched comparisons could help distinguish effects of design revision from differences between participant groups and contexts. Further work should also examine the relationships between didactic evaluation,

usability, presence, immersion, technological self-efficacy, and actual musical learning outcomes. Such studies could clarify not only whether an XR learning design is positively received, but under which pedagogical and institutional conditions it contributes to meaningful musical activity.

V. CONCLUSION

This study examined teachers' evaluations of three XR learning designs and an accompanying professional development format across successive DBR cycles. The findings show generally positive developments, with the clearest and most consistent trends observed for PatchWorld and for the practice relevance, structure, and exchange dimensions of professional development.

The results provide exploratory support for DBR as a framework for developing XR learning designs in music education. Its contribution lies in connecting design, authentic implementation, participant feedback, and revision. The findings further indicate that successful XR integration depends on the alignment of musical objectives, learner activity, pedagogical framing, technological accessibility, and professional development.

Because participant groups differed between workshops and no controlled longitudinal design was used, the observed trends cannot be interpreted causally. Nevertheless, the convergence of construct-level and item-level findings is consistent with positive development across the iterative process, while the present design does not isolate the contribution of design revisions from cohort, location, or other workshop-specific effects.

Future work should examine classroom transfer, long-term adoption, student learning outcomes, and the institutional conditions required for sustainable implementation. More broadly, XR-based innovation in music education should remain centered on musical activity, teachers, and learners rather than on technological novelty.

REFERENCES

- [1] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.
- [2] M. Rossi, "Synaptic sanctum neuroscape: Real-time emotionally adaptive music in a VR immersive environment," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Utrecht, Netherlands, 2024, installation paper. [Online]. Available: https://nime.org/proceedings/2024/nime2024_installations_2.pdf
- [3] S. Rutkowski, M. Adamczyk, A. Pastuła, E. Gos, C. Luque-Moreno, and A. Rutkowska, "Training using a commercial immersive virtual reality system on hand-eye coordination and reaction time in young musicians: A pilot study," *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, p. 1297, 2021. [Online]. Available: <https://doi.org/10.3390/ijerph18031297>
- [4] K. R. J. Keeler, "Video games in music education: The impact of video games on rhythmic performance," *Visions of Research in Music Education*, vol. 37, no. 1, p. Article 5, 2020. [Online]. Available: <https://digitalcommons.lib.uconn.edu/vrme/vol37/iss1/5>
- [5] P. Feneberg and S. Schmidt, "Symphonic spaces: Virtuelle konzerträume aus musikdidaktischer perspektive. ein beitrag zum einsatz und zur gestaltung immersiver lernumgebungen," in *Lehrer:innenbildung und Schulpraxis im internationalen Kontext. Impulse für Wissenstransfer, kulturelle Sensibilität, Sprachbildung und Digitalität*, M. Hackel, W.-K. Giera, J. Buschmann, and T. Fischer, Eds. Münster and New York: Waxmann, 2024, pp. 243–255. [Online]. Available: <https://doi.org/10.31244/9783830998822.21>
- [6] Z. Xuejing and W. A. J. W. Yahaya, "A systematic review of augmented reality in music education," in *Enhancing Learning Experiences With Digital Tools: AI, ChatGPT, and Virtual and Augmented Reality*, M. Sanmugam, B. Edwards, N. Mohd Barkhaya, and Z. Khlaif, Eds. IGI Global Scientific Publishing, 2025, pp. 1–22. [Online]. Available: <https://doi.org/10.4018/979-8-3693-9811-1.ch001>
- [7] P. Feneberg, "Creating music in virtual worlds: Decisions based on visibility and (un-)lucky accidents?" *Journal of Sound, Silence, Image and Technology*, no. 8, pp. 106–126, 2025.
- [8] A. Boem, M. Tomasetti, A. Gabriele, A. Di Scipio, A. Micalizzi, and L. Turchet, "Imagining the musical metaverse: A study on the needs and creative practices of composers and musicians," *Digital Creativity*, vol. 37, no. 2, pp. 161–187, 2026.
- [9] P. Ahner, "Didactic changes in sound compositions with synthesizer apps," *Divergence Press*, 2023. [Online]. Available: <https://doi.org/10.5920/DivP.2023.02>
- [10] —, "Learning environments and learning tasks with synthesiser apps in secondary schools: A design research project," in *Make Music Matter: Music Education Meeting the Needs of Young Learners*, ser. European Perspectives on Music Education, A. Houmann and E. Saether, Eds. Innsbruck, Esslingen, and Bern-Belp: Helbling, 2020, vol. 9, pp. 225–240.
- [11] W. I. Bauer, *Music Learning Today: Digital Pedagogy for Creating, Performing, and Responding to Music*, 2nd ed. New York: Oxford University Press, 2020.
- [12] P. Ahner, M. Ahlers, W. Aigner, S. Chatelain, S. Desmurs, D. Hug, P. Mall, and B. Spieker, "From sheet to beat: Creativity and composition as useful content for digital lifelong learning," in *Liberty – Equity – Creativity: Innovating and Inventing Music in the Classroom*, ser. European Perspectives on Music Education, T. Buchborn, M. Gall, S. Hennessy, and M. Stumpfegger, Eds. Innsbruck: Helbling, 2025, vol. 13, pp. 189–205.
- [13] S. A. Ruthmann, E. S. Tobias, C. Randles, and M. D. Thibeault, "Is it the technology? challenging technological determinism in music education," in *Music Education: Navigating the Future*, C. Randles, Ed. New York: Routledge, 2015, pp. 122–138.
- [14] S. E. McKenney and T. C. Reeves, *Conducting Educational Design Research*, 2nd ed. London: Routledge, 2018. [Online]. Available: <https://doi.org/10.4324/9781315105642>
- [15] L. Bruns and T. Rotsch, "Xr-producing im musikunterricht. didaktisches design im projekt leviko-xr am beispiel „kollaboratives musik erfinden zu mood clips in der xr-applikation patchworld“,“ *Medienpädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, vol. 68, no. XR Fachdidaktik, pp. 171–195, 2026. [Online]. Available: <https://doi.org/10.21240/mpaed/68/2026.01.27.X>
- [16] L. Bruns, B. Saurbier, T. M. Voong, and M. Oehler, "Presence and flow in virtual and mixed realities for music-related educational settings," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 2024, pp. 1–7.
- [17] M. Oehler, L. Bruns, B. Saurbier, and T. M. Voong, "Impact of spatial and communication environments on collaborative music-making in virtual reality: Insights for music education," in *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2025, pp. 619–625.
- [18] M. Oehler, P. Ahner, C. Geiger, L. Bruns, T. Rotsch, S. Dalfovo, T. M. Voong, B. Saurbier, A. T. Holtmann, and J. Feitsch, "Virtual and augmented reality in the classroom: The project LEVIKO-XR," Presented at the 31st European Association for Music in Schools Conference (EAS), Dublin, Ireland, Jun. 2024, 12–15 June 2024. [Online]. Available: <http://dx.doi.org/10.13140/RG.2.2.24662.46402>
- [19] T. Rotsch, S. Dalfovo, P. Ahner, L. Bruns, B. Saurbier, T. M. Voong, M. Oehler, J. Lückhof, M. Bade, and C. Geiger, "Extended reality im musikunterricht. eine handreichung für lehrkräfte," ResearchGate, Jan. 2026, handreichung. [Online]. Available: <https://doi.org/10.13140/RG.2.2.36428.07042>
- [20] T. Rotsch, "Extended reality im musikunterricht. herausforderungen im wissenstransfer," *Diskussion Musikpädagogik*, no. 109, pp. O.26–O.39, 2026, musik lernen:digital (Online).
- [21] V. Frederking, "Von TPACK und DPACK zu SEPACK.digital. ein alternativmodell für fachdidaktisches wissen in der digitalen welt nebst einigen anmerkungen zu blinden flecken und widersprüchen in den KMK-initiativen zur digitalen bildung in deutschland," in *Fachliche Bildung in der digitalen Welt. Digitalisierung, Big Data und KI im Forschungsfokus von 15 Fachdidaktiken*, ser. Allgemeine Fachdidaktik,

- V. Frederking and R. Romeike, Eds. Münster and New York: Waxmann, 2022, vol. 3, pp. 481–522.
- [22] P. Ahner, T. Rotsch, J. Feitsch, J. Lückhof, and L. Bruns, “Rhythmisch-metrische kompetenzen in VR-lernumgebungen – musikdidaktische perspektiven aus dem projekt LEVIKO-XR,” in *Konturen einer postdigitalen Musikpädagogik*, J. Treß, T. Buchborn, and P. Stade, Eds. Olms, in press, in press.
- [23] E. Richter and D. Richter, “Fortbildungsmonitor. ein instrument zur erfassung der prozessqualität von lehrkräftefortbildungen,” Potsdam, 2023. [Online]. Available: <https://doi.org/10.25656/01:27640>
- [24] C. A. Morton, “Music education for “all my relations”,” in *The Oxford Handbook of Philosophy in Music Education*, A. L. Frega and W. D. Bowman, Eds. Oxford: Oxford University Press, 2012, pp. 472–494.
- [25] P. Ahner, “Individuell, vernetzt, motivierend: Learning experience platforms und microlearning an musikhochschulen,” in *Zukunft Musikhochschule!?*, B. Busch and S. Herbst, Eds. Mainz: Schott Verlag, 2026, pp. 263–270.
- [26] S. Seufert, *Innovationsorientiertes Bildungsmanagement: Hochschulentwicklung durch Sicherung der Nachhaltigkeit von E-Learning*. Wiesbaden: VS Verlag für Sozialwissenschaften, 2008. [Online]. Available: <https://doi.org/10.1007/978-3-531-91004-8>
- [27] P. Ahner, M. Bosch Sanfélix, S. Dalfovo, T. Rotsch, and L. Werner, “Hochschulen für musik in postdigitalen transformationen,” *Diskussion Musikpädagogik*, no. 109, pp. O.18–O.25, 2026, musik lernen:digital, Online-Beitrag Nr. 3. [Online]. Available: <https://www.junker-verlag.com/dmp-109-o-c>